Automatic Facial Expression Tracking for 4D Range Scans

Guofu Xiang, Xiangyang Ju, and Patrik O'B. Holt
Cognitive Engineering Research Group, School of Computing, The Robert Gordon University,
Aberdeen, AB25 1HG, UK

Abstract— *This paper presents a fully automatic approach of spatio-temporal facial expression tracking for 4D range scans without any manual interventions (such as specifying landmarks). The approach consists of three steps: rigid registration, facial model reconstruction, and facial expression tracking. A Scaling Iterative Closest Points (SICP) algorithm is introduced to compute the optimal rigid registration between a template facial model and a range scan with consideration of the scale problem. A deformable model, physically based on thin shells, is proposed to faithfully reconstruct the facial surface and texture from that range data. And then the reconstructed facial model is used to track facial expressions presented in a sequence of range scans by the deformable model.*

Keywords: Scaling Iterative Closest Points (SICP) algorithm, deformable model, template fitting, range scan, face tracking

1. Introduction

Modeling and animating realistic facial models is a substantial challenge in computer graphics, especially for facial expressions, because we are so familiar with human faces and very sensitive to "unnatural" subtle changes in faces. Such a challenge has drawn intensive academic and industrial interest in this area [1], [2]. Creating a convincing facial animation requires a tremendous amount of artistry and manual work. One way to reduce such painstaking work is to automatically capture facial motions directly from real human faces and then use the captured motion data to drive virtual characters to perform the similar facial expressions. The key problems to such an automatic animation framework are facial model reconstruction and facial expression tracking.

Recent advances in 3D scanning technology [3], [4], [5] enable us to easily acquire a time sequence of 3D range scans of human faces performing various facial expressions. However, the raw range scans are usually too noisy and incomplete to suitable for further analysis and animation. Moreover, there is lack of correspondences among the sequences. Thus, template-fitting methods are widely used, with the intention of filling holes, reducing the noise level, capturing characteristic features of range scans [3], [4], [6] and establishing dense point correspondences. Some manual interventions are generally required during the template fitting to provide a small set of feature landmarks in order to roughly warp the template to range scans [3], [4], [6]. However, manually positioning of landmarks seems to be tedious and error-prone.

In this paper, we present a *fully automatic* approach to reconstructing 3D facial models and textures from range scans without requiring manual interventions. This paper makes several specific technical contributions. First, we introduce a *Scaling Iterative Closest Points* (SICP) algorithm to compute the optimal rigid registrations between a generic template facial model and range scans with the consideration of the scale problem. Second, we propose a deformable model to reconstruct facial models and textures from range scans and we also use the deformable model to track facial expressions presented in 4D range scans.

In the following section, we review some topics related to our work. In Section 3, we present the details of SICP to rigidly register a template facial model to range scans with different scales and show how to use our deformable model to reconstruct facial models and textures from range scans, and to track facial expressions. Results and conclusions are presented in Sections 4 and 6, respectively.

2. Related Work

Modeling and tracking faces is an active research field in computer graphics and computer vision. Here we review three topics most related to our current work: ICP-based registration, template fitting, face tracking. Other related work is discussed throughout the paper, as appropriate.

a) ICP-based Registration: Since the first paper of ICP [7], ICP has been widely used for geometric alignment of 3D models and many variants of ICP have been proposed [8]. Generally, the original ICP can only deal with models with the same scale. To account for the scale problem, Du *et al.* proposed an Iterative Closest Point with Bounded Scale (ICPBS) algorithm which integrated a scale parameter with boundaries into the original ICP algorithm [9], but it's unclear how to determine the upper and lower boundaries of scales that contain the optimal scale.

b) Template Fitting: Due to its great challenge in many research fields, numerous research efforts are devoted to establishing correspondences between different meshes [10]. The template-fitting method [6], [11] deforms a template to a target object to minimize the combining errors of smoothness and fitness between them. Recently, template fitting has become particular popular due to its simplicity and robustness to noisy range data [12], [4]. Our reconstruction method shares the similar idea, but it is derived from physically-based elastic deformations of thin shells by linear the variational deformation technique [13].

c) Face Tracking: There is a vast body of work on tracking the human face, with applications ranging form motion capture to human-computer interaction. Zhang *et al.* [3] presented an automatic offline face tracking method on 3D range data. The resulting facial expression sequences are then used in an interactive face modeling and animation application. Weise *et al.* [4] proposed real-time facial expression tracking with transfer to another person's face. All these methods require manually labeled feature points for initial warping of the generic template to range scans.

In contrast, our method does not require any manual interventions. We solve the initial rigid registration using SICP, which takes the consideration of the scale problem. Our deformable model is physically based on thin shells and facial model and texture reconstructions are solved in the same framework.

3. Automatic Facial Expression Tracking

In this paper, we assumed that the sequences of range scans to track were upright front faces with limited rotation, in which some other unwanted parts (such as neck, shoulder) might also present. Given such range scan sequences, our goal is to build a new facial model with texture to reflect the shape and texture of the range scans from a template facial model and then track the facial expressions presented in the range scans.

Our facial model reconstruction method consists of two steps: the first step is to compute the initial rigid registration between a template and a range scan; the second step is to iteratively deform the template model toward the range scan to capture the shape of the range scan and the texture is obtained in a similar way.

We preferred triangle meshes for the representation of our models and range scans for efficiency and simplicity. Before elaborating our method, let us introduce some notations used in this paper. A triangle mesh $\mathcal{M}$ consists of a geometrical and a topological component, i.e., $\mathcal{M} = (\mathcal{P}, \mathcal{K})$, where the latter can be represented by a simplicial complex with a set of vertices $\mathcal{V} = \{v_i, 1 \leq i \leq |\mathcal{V}|\}$[1], edges $\mathcal{E} = \{e_i \in \mathcal{V}\times\mathcal{V}, 1 \leq i \leq |\mathcal{E}|\}$ and triangles $\mathcal{F} = \{f_i \in \mathcal{V}\times\mathcal{V}\times\mathcal{V}, 1 \leq i \leq |\mathcal{F}|\}$. The geometric embedding of a triangle mesh into $\mathbb{R}^3$ is specified by associating a 3D position $\mathbf{p}_i$ for each vertex $v_i \in \mathcal{V}$: $\mathcal{P} = \{\mathbf{p}_i := \mathbf{p}(v_i) \in \mathbb{R}^3, 1 \leq i \leq |\mathcal{V}|\}$.

[1] $|\cdot|$ denotes the number of elements in the set.

3.1 SICP Registration

In order to reconstruct the surface of a range scan using a template, we need first roughly place the template close to the range scan. Traditionally, this is done by manually specifying a small set of landmarks [3], [4], [6], [11]. Our method deals with this problem with no manual intervention.

Since the template facial model and the range scans of human faces have much similarity in shape, it is intuitive to use the ICP algorithm to compute the initial rigid registrations between them. However, there is a challenge dealing with the scale problem, because the size of the facial region in the range scans is not known a priori and the range scans may also include some unwanted parts.

To deal with the scale problem, we employed an extension version of the ICP algorithm called the Scaling Iterative Closest Points (SICP) algorithm [14], which integrates a scale parameter s to the original ICP equation and iteratively refines the scale from an estimated initial scale until convergence.

Given a template mesh $\mathcal{M}_{\text{template}}$ and a range scan mesh $\mathcal{M}_{\text{scan}}$, the goal of SICP is to find the transformation (scale s, rotation $\mathbf{R} \in \mathbb{R}^{3\times 3}$ and translation $\mathbf{t} \in \mathbb{R}^3$) so that the distance between the registered template mesh $\mathcal{M}'_{\text{template}}$ and $\mathcal{M}_{\text{scan}}$ is as close as possible. Obviously, we should avoid degenerate cases such as $s = 0$ by providing a good initial value for s.

As the original ICP algorithm, SICP is an iterative algorithm, which iteratively refines the registration based on previous registrations until it satisfies a certain termination condition. Let us denote the series of registrations by $\mathcal{T} = \{\mathbf{T}_k = (s_k, \mathbf{R}_k, \mathbf{t}_k), 0 \leq k \leq |\mathcal{T}|\}$. Then the registration process can be formulated mathematically as follows,

$$\mathcal{C}_{k+1} = \{\arg \min_{c \in \mathcal{M}_{\text{scan}}} d(s_k \mathbf{R}_k \mathbf{p}_i + \mathbf{t}_k, \mathbf{c})\}\,, \tag{1}$$

$$(s_{k+1}, \mathbf{R}_{k+1}, \mathbf{t}_{k+1}) = \arg \min_{s,\mathbf{R},\mathbf{t}} \sum_{i=1}^{|\mathcal{P}_{\text{template}}|} \|s\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{c}_i\|^2, \mathbf{c}_i \in \mathcal{C}_k\,, \tag{2}$$

where $\mathbf{p}_i \in \mathcal{M}_{\text{template}}$, $d(\cdot)$ is a distance function. Equation 1 is to find the corresponding closest points on $\mathcal{M}_{\text{scan}}$ for the points of $\mathcal{M}_{\text{template}}$ and Equation 2 is the absolute orientation problem [15].

As mentioned above, the initial registration state, $s_0, \mathbf{R}_0, \mathbf{t}_0$, is important for the local convergence of SICP. In our examples, we set the initial values as following,

$$s_0 = \frac{\sum_{i=0}^{N} |\mathbf{q}_i - \bar{\mathbf{q}}|/N}{\sum_{i=0}^{M} |\mathbf{p}_i - \bar{\mathbf{p}}|/M}, \quad \mathbf{R}_0 = \mathbf{I}, \quad \mathbf{t}_0 = \bar{\mathbf{q}} - s_0 \mathbf{R}_0 \bar{\mathbf{p}}\,, \tag{3}$$

where $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$ are the centroids of the template and the scan meshes, M and N the number of points of the two meshes, and $\mathbf{I}$ the 3×3 identity matrix. Although SICP has many degenerate cases and does not guarantee the global convergence, our tests show its capability to register the template to different range scans (see Figures 1).

3.2 Deformable Model

Due to the shape diversities between the template facial model and range scans, we need further deform the template after the initial rigid registration. There are two criteria that should be considered during the deformation process. One is the regularity that penalizes dramatic changes in mesh. Another criterion is the fitting error, which can be formulated as the total distance between corresponding points.

Since the template mesh is a two-manifold surface, the change of the surface can be measured by the change of the first and the second fundamental forms and therefore yields a measure of stretching and bending. Given a two-manifold surface $\mathcal{S}$, after deformation, it becomes $\mathcal{S}'$, we can represent the deformed surface $\mathcal{S}'$ by $\mathbf{p}' = \mathbf{p} + \mathbf{d}$, where $\mathbf{p} \in \mathcal{S}$, $\mathbf{p}' \in \mathcal{S}'$, and $\mathbf{d}$ is the displacement. The minimization of the physically-based elastic energies yields the so-called Euler-Lagrange partial differential equation (PDE) [13]:

$$-k_s \Delta \mathbf{d} + k_b \Delta^2 \mathbf{d} = 0\,, \tag{4}$$

where k_s and k_b are coefficients, Δ and Δ^2 represent the Laplacian and the bi-Laplacian operator, respectively. The Laplacian operator can be extended to triangle meshes to obtain the discrete form of the Laplace-Beltrami operator $\Delta_{\mathcal{M}}$ (refer to [13]). Thus, we can formulate our deformable model as follows,

$$\begin{aligned} \min_{\mathbf{d}_i} \sum_{i=1}^{M} \| - k_s \Delta_{\mathcal{M}_{\text{template}}} \mathbf{d}_i + k_b \Delta^2_{\mathcal{M}_{\text{template}}} \mathbf{d}_i \|^2 \\ + k_c \sum_{i=1}^{M} w_i \| \mathbf{d}_i - (\mathbf{c}_i - \mathbf{p}_i) \|^2\,, \end{aligned} \tag{5}$$

where $\mathbf{p}_i \in \mathcal{P}_{\text{template}}$, $\mathbf{c}_i \in \mathcal{M}_{\text{scan}}$ is the corresponding closest point of $\mathbf{p}_i$, $\mathbf{d}_i$ is the unknown displacement, and k_s, k_b, k_c represent the contribution of stretching, bending and fitting in the total energy, respectively. $w_i = 1$ if the corresponding closest point satisfies a certain compatible conditions, otherwise 0. We employed the similar compatible conditions as [11], [16] to reject pseudo point matching, such as, requiring the angle between two corresponding normals should be greater than 60 degrees, rejecting boundary vertices. The minimization problem can be reformulated as a sparse linear system in term of least squares [13].

An annealing-like deformation scheme is employed in our experiments. At the initial stage, k_s and k_b are set to relatively large values compared to k_c (In our tests, k_s, k_b and k_c are initially set to 50, 20, 2, respectively). Because at the initial stage we cannot estimate good correspondences between the template and the range scan by the closest points due to the shape diversity and large values of k_s and k_b do not allow dramatic change of the mesh. Then we relax the stiffness of the template facial model by gradually decreasing the values of k_s and k_b toward 1.

3.3 Texture Reconstruction

Texture can improve the reality of facial models. Thus it is desirable to make the textures available for the reconstructed facial models. However, the original range scans usually have holes (missing data). We cannot find all the texture coordinates for the reconstructed facial models.

We solve the texture reconstruction problem in the similar way proposed in the previous section, but here we consider the texture coordinates $\mathbf{u}_i \in \mathbb{R}^2$ as the unknown variables and the equation becomes

$$\begin{aligned} \min_{\mathbf{u}_i} \sum_{i=1}^{M} \| - k_s \Delta_{\mathcal{M}_{\text{template}}} \mathbf{u}_i + k_u \Delta^2_{\mathcal{M}_{\text{template}}} \mathbf{u}_i \|^2 \\ + k_c \sum_{i=1}^{M} w_i \| \mathbf{u}_i - \mathbf{u}'_i \|^2, \end{aligned} \tag{6}$$

where $\mathbf{u}'_i$ is the texture coordinates of the corresponding closest point on the range scan for the point $\mathbf{p}_i$.

When reformulating Equations 5 and 6 in matrix form, we can see that the two equations have the same sparse matrix and only differ in the right hand side. Thus the texture reconstruction can be efficiently solved because the sparse matrix is only factorized once.

3.4 Facial Expression Tracking

In this section, we present the procedure of facial expression tracking. Given a sequences of range scans of a human face performing facial expressions, $\mathcal{S} = \{\mathcal{M}^t_{\text{scan}}, t = 0, \ldots, n\}$, without lost of generality, we denote by $\mathcal{M}^0_{\text{scan}}$ the reference neutral range scan. The template facial model $\mathcal{M}_{\text{template}}$ is first registered to $\mathcal{M}^0_{\text{scan}}$ using SICP and the aligned template model $\bar{\mathcal{M}}^0_{\text{template}}$ is obtained. Then the deformable model is used to non-rigidly register the initial aligned template $\bar{\mathcal{M}}^0_{\text{template}}$ to $\mathcal{M}^0_{\text{scan}}$, yielding the reconstructed facial model $\mathcal{M}^0_{\text{template}}$ for that range scan. For the subsequent range scans, we use the previous deformed template $\mathcal{M}^t_{\text{template}}$ to non-rigidly register to $\mathcal{M}^{t+1}_{\text{scan}}$. This tracking procedure can be described in Equation 7.

$$\begin{array}{ccc} \mathcal{M}^t_{\text{scan}} & & \mathcal{M}^{t+1}_{\text{scan}} \\ \downarrow & & \downarrow \\ \mathcal{M}^t_{\text{template}} & \longrightarrow & \mathcal{M}^{t+1}_{\text{template}} \end{array} \tag{7}$$

4. Results

We tested our method on two data sets (a male and a female) from the 4D facial expression database [5], which collected the spatio-temporal range scan sequences of six facial expressions (angry, disgust, fear, happy, sad, surprise) for each person. The sequences of range scans are captured

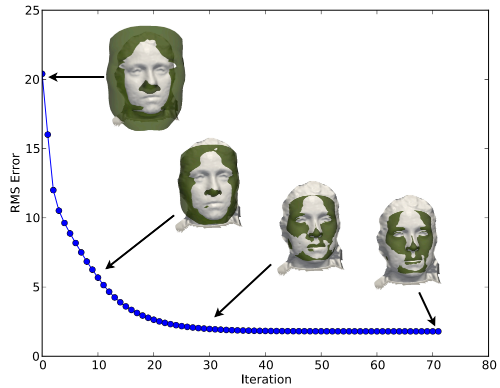

Fig. 1: The RMS error of SICP rigid registration. The inset figures show the overlap between the template model and the range scan during the rigid registration.

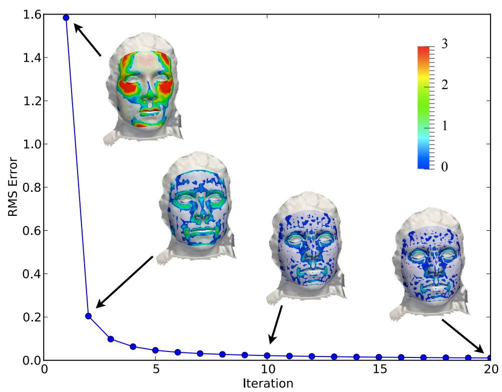

Fig. 2: Deformation process of the deformable model. The color mapping shows the distances between the template and the range scan.

at a speed of 25 frames per second. Each facial expression lasts about 4 seconds, thus there are about 100 frames of range scans for each expression.

Figure 1 shows the curve of the root-mean-squared (RMS) error during the SICP registration of the template to the range scan. The curve definitely indicates the convergence of SICP, which is also shown by the inset figures.

Figure 2 shows the reconstruction error during the facial model reconstruction. The curve presents the average RMS distance from the template to the range scan and the distribution of the reconstruction error is shown the color mapping inset figures. From this figure, we can see that the reconstruction error rapidly decreases across the face region in the first several iteration steps.

The facial expression tracking results are shown in Figure 3. The first columns in Figure 3 (a) and (b) are the range scans and the template facial model and the rest columns show the various facial expressions, which are presented in the range scan sequences, and their corresponding tracked expressions in the template facial models. The results show that our method is able to reconstruct new facial models from range scans with their textures and it can also track the facial expressions expect for that with very large deformation.

5. Limitations

There are some limitations existing in our method. First, currently our method only employs the closest point constraints, which restricts it for tracking large and fast motion, especially in the chin region (see the surprise expressions in Figure 3). The motion of the chin often exhibits fast and abrupt and hence our deformable can fail to track it correctly. Second, our method is based on the assumption that the acquisition rate is very high so that the change of facial motion between two consecutive frames is not very large. Third, we haven't integrated any texture information into our deformable model. Actually, the textures contain very rich information about the facial motion, which can be detected using the optical flow technique [17]. Adding such constraints into our method will improve the accuracy of the facial expression tracking.

6. Conclusion

We have presented an automatic algorithm for 3D facial model and texture reconstruction from range scans of human faces. The proposed deformable model is also able to track facial expressions presented in 4D range scans if the facial motion is not very fast and large. One of the main benefits of our method is fully automatic. Our method requires no manual intervention and we do not require a small set of corresponding feature landmarks. Key to the success of our algorithm is the robust rigid registration based on Scaling Iterative Closest Points (SICP) algorithm and the template fitting based on the proposed deformable model.

As future work, we plan to employ the optical flow method to integrate the texture information into the deformable to improve the accuracy and of facial expression tracking and extend it to account for fast motion.

Acknowledgment

This work is supported by the Robert Gordon University Research Development Initiative (RDI) PhD Studentship. The authors would like to thank the organizers of the Facial Expression Database (FED) [5] for providing 4D range data.

References

[1] J. Haber and D. Terzopoulos, "Facial modeling and animation," in *SIGGRAPH '04: ACM SIGGRAPH 2004 Courses*. Los Angeles, California, USA: ACM, Aug. 8–12 2004, p. 6.

[2] F. Pighin and J. P. Lewis, "Performance-driven facial animation," in *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses*. San Diego, California, USA: ACM, July 30–Aug. 3 2006.

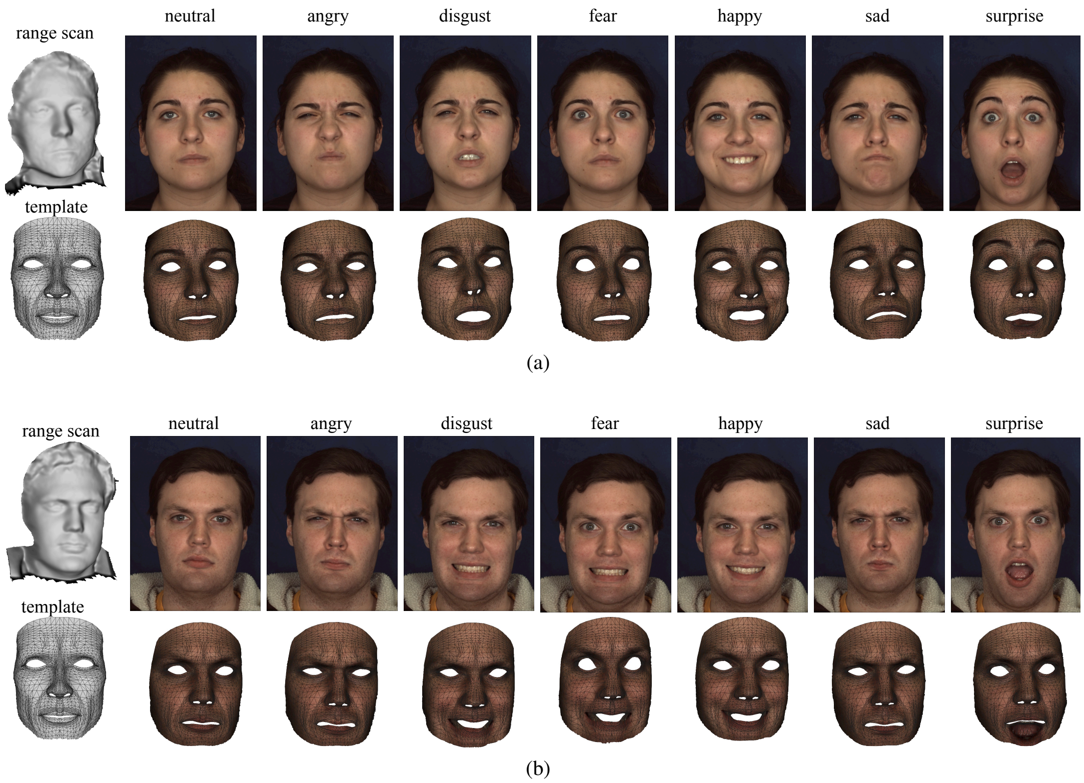

Fig. 3: The results of facial expression tracking. The two sequences of range scans with facial expressions are from the facial expression database [5]. The sequences are recorded at the rate of 25 fps. Each expression lasts about 4 seconds. The first columns in (a) and (b) show the 3D range scan and the template facial model. The reset columns are the facial expressions extracted from the sequences and the corresponding tracked expressions in the template facial model.

[3] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz, "Spacetime faces: high resolution capture for modeling and animation," in *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*. Los Angeles, California, USA: ACM, Aug. 8–12 2004, pp. 548–558.

[4] T. Weise, H. Li, L. Van Gool, and M. Pauly, "Face/off: live facial puppetry," in *SCA '09: Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. New Orleans, Louisiana: ACM, Aug. 1–2 2009, pp. 7–16.

[5] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Rale, "A high-resolution 3D dynamical facial expression database," in *FGR '08: Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*. Amsterdam, The Netherlands: IEEE Computer Society, Sept. 17–19 2008, pp. 1–6.

[6] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: reconstruction and parameterization from range scans," in *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*. San Diego, California, USA: ACM, July 27–31 2003, pp. 587–594.

[7] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," vol. 14, no. 2, pp. 239–256, Feb. 1992.

[8] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *3DIM '01: Proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling*. Quebec City, Canada: IEEE Computer Society, May 28–June 1 2001, pp. 145–152.

[9] S. Du, N. Zheng, S. Ying, and J. Wei, "ICP with bounded scale for registration of m-D point sets," in *ICME '07: Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*. Beijing, China: IEEE Computer Society, July 2–5 2007, pp. 1291–1294.

[10] K. Hormann, K. Polthier, and A. Sheffer, "Mesh parameterization: theory and practice," in *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 Courses*. Singapore: ACM, Dec. 10–13 2008, pp. 1–87.

[11] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," in *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*. Los Angeles, California, USA: ACM, Aug. 8–12 2004, pp. 399–405.

[12] H. Li, B. Adams, L. J. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," in *SIGGRAPH Asia '09: ACM SIGGRAPH Asia 2009 Papers*. New York, NY, USA: ACM, Sept. 16–19 2009, pp. 175:1–10.

[13] M. Botsch and O. Sorkine, "On linear variational surface deformation methods," vol. 14, no. 1, pp. 213–230, Jan./Feb. 2008.

[14] S. Du, N. Zheng, S. Ying, Q. You, and Y. Wu, "An extension of the ICP algorithm considering scale factor," in *ICIP '07: Proceedings of the International Conference on Image Processing*, vol. 5. San Antonio, Texas, USA: IEEE Computer Society, Sept. 16–19 2007, pp. 193–196.

[15] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," vol. 4, no. 4, pp. 629–642, Apr. 1987.

[16] G. Turk and M. Levoy, "Zippered polygon meshes from range images," in *SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. Orlando, Florida, USA: ACM, July 24–29 1994, pp. 311–318.

[17] D. DeCarlo and D. Metaxs, "Optical flow constraints on deformable models with applications to face tracking," *International Journal of Computer Vision*, vol. 38, no. 2, pp. 99–127, July 2000.